The recoding of DNA sequences to enable them to be expressed in yeasts, and the transformed yeasts obtained

The present invention relates to the recoding of DNA sequences which encode proteins which contain regions having a high content of codons which are poorly translated by yeasts, in particular which encode proteins of plant origin, such as the P450 cytochromes of plant origin, and to their expression in yeasts.

It is known that certain sequences encoding proteins of interest, in particular proteins of plant origin, are not readily translated in yeasts. This applies, in particular, to proteins which possess regions having a high content of codons which are poorly suited to yeasts, in particular leucine codons, such as some P450 cytochromes of plant origin. Some systems which have been developed for improving the expression of P450 cytochromes of animal or plant origin in yeasts, such as those described by Pompon et al. (*Methods Enzymol.*, 272, 1996, 51-64; WO 97/10344), have turned out to be unsuitable for large numbers of P450 cytochromes which encompass regions having a high content of codons which are poorly suited to yeasts.

The P450 cytochromes constitute a superfamily of membrane enzymes of the monooxygenase type which are able to oxidize a large family of generally hydrophobic substrates. The reactions are most frequently characterized by the oxidation of C-H or C=C bonds, and

of heteroatoms, and, more rarely, by the reduction of nitro groups or by dehalogenation. More specifically, these enzymes are involved in the metabolism of xenobiotic substances and drugs and in the biosynthesis of secondary metabolites in plants, some of which have organoleptic or pharmacodynamic properties.

As a consequence, the P450 cytochromes are used, in particular, in:

- the *in vitro* diagnosis of the formation of toxic or mutagenic metabolites (molecules of natural origin, pollutants, drugs, pesticides, etc.), making it possible, in particular, to develop novel active molecules (pharmaceutical, agrochemistry),
- the identification and destruction of molecules which are toxic for, or pollute, the environment,
- the enzymic synthesis of novel molecules.

The search for heterologous expression of P450 cytochromes by host cells, more specifically yeasts, is therefore important for obtaining controlled production of this enzyme in large quantity, either for isolating it and using it in the above-listed processes, or for using the transformed cells directly for the said processes without previously isolating the enzyme.

The present invention provides a solution to the abovementioned problem, enabling proteins which contain regions having a high content of codons which

are poorly suited to yeasts, in particular P450 cytochromes of plant origin, to be expressed in yeasts.

The present invention therefore relates to a DNA sequence, in particular a cDNA sequence, which encodes a protein of interest which contains regions having a high content of codons which are poorly suited to yeasts, characterized in that a sufficient number of codons which are poorly suited to yeasts is replaced with corresponding codons which are well-suited to yeasts in the said regions having a high content of codons which are poorly suited to yeasts.

Within the meaning of the present invention, "codons which are poorly suited to yeasts" are understood as being codons whose frequency of use by yeasts is less than or equal to approximately 13 per 1000, preferably less than or equal to approximately 12 per 1000, more preferably less than or equal to approximately 10 per 1000. The frequency at which codons are used by yeasts, more specifically by *S. cerevisiae*, is described, in particular, in "Codon usage data base from Yasukazu Nakamura" (http://www.dna.affrc.go.jp/~nakamura/codon.html). This applies, in particular, to codons CTC, CTG and CTT, which encode leucine, to codons CGG, CGC, CGA, CGT and AGG, which encode arginine, to codons GCG and GCC, which encode alanine, to codons GGG, GGC and GGA, which encode glycine, and to codons CCG and CCC, which encode proline. The codons which are poorly suited to yeasts

SUB B' CON't.

in accordance with the invention are, more specifically, codons CTC and CTG, which encode leucine, CGG, CGC, CGA, CGT and AGG, which encode arginine, codons GCG and GCC, which encode alanine, GGG and GGC, which encode glycine, and codons CCG and CCC, which encode proline.

Within the meaning of the present invention, "corresponding codons which are well-suited to yeasts" are understood as being the codons which correspond to the codons which are poorly suited to yeasts and which encode the same amino acids, and whose frequency of use by yeasts is greater than 15 per 1000, preferably greater than or equal to 18 per 1000, more preferably greater than or equal to 20 per 1000. This applies, in particular, to codons TTG and TTA, preferably TTG, which encode leucine, to codon AGA, which encodes arginine, to codons GCT and GCA, preferably GCT, which encode alanine, to codon GGT, which encodes glycine, and to codon CCA, which encodes proline.

Within the meaning of the present invention, "region having a high content of codons which are poorly suited to yeasts" is understood as being any region of the DNA sequence which contains at least 2 poorly suited codons among 10 consecutive codons, with it being possible for the two codons to be adjacent or separated by up to 8 other codons. According to one preferred embodiment of the invention, the regions having a high content of poorly suited codons contain

2, 3, 4, 5 or 6 poorly suited codons per 10 consecutive codons, or contain at least 2 or 3 adjacent poorly suited codons.

Within the meaning of the present invention, "sufficient number of codons" is understood as being the number of codons which it is necessary and sufficient to replace in order to observe a substantial improvement in their expression in yeasts. Advantageously, at least 50% of the codons which are poorly suited to yeasts in the high-content region under consideration are replaced with well-suited codons. Preferably, at least 75% of the poorly suited codons of the said region are replaced, with 100% of the poorly suited codons more preferably being replaced.

Within the meaning of the present invention, "substantial improvement" is understood as being either a detectable expression when no expression of the reference sequence is observed, or an increase in expression as compared with the level at which the reference sequence is expressed.

Within the meaning of the present invention, "reference sequence" designates any sequence which encodes a protein of interest and which is modified in accordance with the invention in order to promote its expression in yeasts.

The present invention is particularly well suited to DNA sequences, in particular cDNA sequences,

which encode proteins of interest which contain regions having a high content of leucine and in which a sufficient number of CTC codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons, or in which a sufficient number of CTC and CTG codons encoding leucine in the said region having a high content of leucine is replaced with TTG and/or TTA codons, preferably with a TTG codon.

Within the meaning of the present invention, "region having a high content of leucine" is understood as being a region which contains at least 2 leucines among 10 consecutive amino acids in the protein of interest, with it being possible for the two leucines to be adjacent or separated by up to 8 other amino acids. According to one preferred embodiment of the invention, the regions having a high content of leucine contain 2, 3, 4, 5 or 6 leucines per 10 consecutive amino acids, or contain at least 2 or 3 adjacent leucines.

According to a preferred embodiment of the invention, at least 50% of the CTC or CTC and CTG codons of the region having a high content of leucine are replaced with TTG or TTA codons, with at least 75% of the CTC or CTC and CTG codons of the said region preferably being replaced, and 100% of the CTC or CTC and CTG codons more preferably being replaced.

Advantageously, the present invention is

particularly suitable for DNA sequences whose general content of poorly suited codons is at least 20%, more preferably at least 30%, as compared with the total number of codons in the reference sequence.

Advantageously, when the reference sequence contains at least one 5′ region having a high content of poorly suited codons, the recoding of this 5′ region alone makes it possible to obtain a substantial improvement in the expression of the protein of interest in yeasts. The length of the 5′ region to be recoded in accordance with the invention will vary depending on the length of the region having a high content of poorly suited codons. This length will advantageously be at least four codons, in particular when this region contains at least two adjacent poor codons, up to approximately 40 codons or more.

However, it is not necessary, according to the invention, to recode all the reference sequence, but only the regions having a high content of poor codons, in particular the 5′ region on its own, in order to obtain a substantial improvement in the expression of the protein of interest in yeasts.

Advantageously, the DNA sequence encoding a protein of interest is an isolated DNA sequence of natural origin, in particular of plant origin. The invention is particularly advantageous for sequences which originate from monocotyledonous or dicotyledonous plants, preferably monocotyledonous plants, in

particular of the graminae family, such as wheat, barley, oats, rice, maize, sorghum, cane sugar, etc.

According to a preferred embodiment of the invention, the DNA sequence encodes an enzyme, in particular a cytochrome P450, which is preferably of plant origin. These P450 cytochromes exhibit a high content of poorly suited codons, in particular encoding leucine, in their N-terminal region; it is in the 5'-terminal coding region that the poorly suited codons are replaced.

The present invention also relates to a chimeric gene which comprises a DNA sequence which has been modified as above and heterologous 5' and 3' regulatory elements which are able to function in a yeast, that is to say which are able to control the expression of the protein of interest in the yeast. Such regulatory elements are well known to the skilled person and are described, in particular, by Rozman et al. (Genomics, 38, 1996, 371-381) and by Nacken et al. (Gene, 175, 1996, 253-260, *Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae*).

The present invention also relates to a vector for transforming yeasts which contains at least one chimeric gene as described above. It also relates to a process for transforming yeasts with the said vector and to the transformed yeasts which are obtained. It finally relates to a process for producing

a heterologous protein of interest in a transformed yeast, with the sequence which encodes the said protein of interest being such as defined above.

The process for producing a heterologous protein of interest in a transformed yeast comprises the steps of:

a) transforming a yeast with a vector which is able to replicate in yeasts and which contains a modified DNA sequence as defined above and heterologous 5′ and 3′ regulatory elements which are able to function in a yeast,

b) culturing the transformed yeast, and

c) extracting the protein of interest from the yeast culture.

When the protein of interest is an enzyme which is suitable for transforming a substrate, such as a cytochrome P450, the enzyme which has been extracted from the yeast culture is then used for catalysing the transformation of the said substrate.

However, the catalysis can be carried out, without requiring the extraction of the yeast, by culturing the transformed yeast in the presence of the said substrate.

The present invention also relates, therefore, to a process for transforming a substrate by enzymic catalysis using an enzyme which is expressed in a yeast, which process comprises the steps of

a) culturing the yeast which has been

transformed in accordance with the invention in the presence of the substrate to be transformed, then

b) recovering the transformed substrate from the yeast culture.

When the yeast has been transformed for expressing a cytochrome P450, the reaction which is catalysed by the enzyme is an oxidation reaction, more specifically a reaction in which C-H or C=C bonds are oxidized.

The techniques for transforming and culturing yeasts are known to the skilled person, and are described, for example, in *Methods in Enzymology* (Vol. 194, 1991).

Yeasts which are of use in accordance with the invention are selected, in particular, from the genera *Saccharomyces*, *Kluyveromyces*, *Hansenula*, *Pichia* and *Yarrowia*. Advantageously, the yeast belongs to the *Saccaromyces* genus, and is in particular *S. cerevisiae*.

Other characteristics of the invention will become apparent in the light of the examples which follow.

Example 1: Production of a wheat cDNA gene library, and identification of the CYP73A17 sequence

The wheat cytochrome P450 CYP73A17 sequence was obtained by screening a young wheat plantlet (shoots and roots without the caryopses) cDNA library which was constructed in the vector λ-ZapII (Stratagene) in accordance with the supplier's

instructions.

1. Production of the cDNA library

Triticum aestivum (L. cv. Darius) seeds which had been coated with cloquintocet-mexyl (0.1% per dry weight of seed) are cultured in plastic boxes on two layers of damp gauze until shoots having a size of 3 to 5 mm are obtained. The water in the boxes is then replaced with a solution of 4 mM sodium phenobarbital and the wheat is cultured until the shoots are approximately 1 cm in size.

The cDNA library is constructed in the λ-ZapII (Stratagene) vector, in accordance with the supplier's protocol and instructions, using 5 μg of poly(A)$^+$ RNA (Lesot, A., Benveniste, I., Hasenfratz, M.P., Durst, F. (1990) Induction of NADPH cytochrome P450(c) reductase in wounded tissues from *Helianthus tuberosus* tubers. Plant Cell Physiol., 31, 1177-1182) which were isolated from the treated roots and shoots.

2. Screening the cDNA library

5×10^5 lysis plaques from the previously obtained λ-ZapII library are screened using a probe which corresponds to the complete coding sequence of *Helianthus tuberosus* CYP73A1, and which has been labelled by random priming with [α-^{32}P]dCTP. The filters are prehybridized and hybridized at low stringency at 55°C in accordance with the standard protocols. The membranes are washed twice for 10 minutes with 2 × SSC, 0.1% SDS, and once for 10 minutes with 0.2 × SSC, 0.1%

SDS at ambient temperature, then twice for 30 minutes with 0.2 x SSC, 0.1% SDS at 45°C. The inserts of the positive lysis plaques are analysed by PCR (polymerization chain reaction) and hybridization in order to determine their size. The clones containing inserts which hybridize with CYP73A1 under the above-described conditions and which are greater than 1.5 kbp in size are rescreened before excision of the pBluescript plasmid in accordance with the supplier's (Stratagene) protocol and sequencing using the Ready Reaction Dye Deoxy Terminator Cycle prism technique developed by Applied Biosystems Inc. A full length clone is then identified by alignment with CYP73A1.

The wheat cytochrome P450 CYP73A17 which is encoded by the isolated sequence (~~sequence identifier No. 1~~) exhibits 76.2% identity with the *Helianthus tuberosus* CYP73A1.

Example 2: Alterations to the sequence encoding the wheat cytochrome P450 CYP73A17

Contrary to the situation with regard to *Helianthus tuberosus* CYP73A1, which can be expressed in yeasts (Urban et al., 1994), repeated attempts to express wheat CYP73A17 in yeasts using the same customary techniques proved to be fruitless when the nucleotide sequence was not altered at the time it was inserted into the expression vector (verification by sequencing). No protein is detected by spectrophotometry or by immunoblotting, just as no

enzymic activity is detectable in the microsomes of transformed and induced yeast.

1. Alteration of the coding sequence

The sequence encoding wheat CYP73A17 (SEQ. ID No. 1) was therefore altered, in three different ways, by PCR-induced mutagenesis, as follows:

The *BamHI* and *EcoRI* restriction sites were respectively introduced by PCR just upstream of the ATG codon and just downstream of the stop codon of the CYP73A17 coding sequence (source, origin) using the sense and reverse primers described below, with the restriction sites being *BamHI* in the case of the sense primers Rec1 (SEQ ID No. 3), Rec2 (SEQ ID No. 4) and Rec3 (SEQ ID No. 5), and *EcoRI* in the case of the reverse primer (SEQ ID No. 6).

A primer, represented by SEQ ID No. 2, was also employed for enabling yeasts to be transformed with the unmodified (native) sequence encoding wheat CYP73A17.

The five primers described above were obtained from Eurogentech, and were synthesized and purified in accordance with customary methods.

For each alteration using the four different sense primers, the mode of operation is as follows:

The reaction mixture (20 mM Tris-HCl, pH 8.75, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X100, 0.1 mg/ml BSA, 5% (v/v) DMSO, 300 μM dNTP, 20 pmoles of each primer, 150 ng of template, total

volume 50 µl) is preheated at 94°C for 2 minutes before adding 5 units of Pfu DNA polymerase (Stratagene). After 2 minutes at 94°C, 30 amplification cycles are carried out as follows: 1 minute of denaturation at 94°C, 2 minutes of hybridization at 55°C, 2 minutes of extension at 72°C. The reaction is completed by 10 minutes of extension at 72°C.

For each primer, a sequence is obtained which is derived from sequence ID No. 1, and which is represented, in the case of the altered coding sequences, by the sequences ID No. 7, No. 8 and No. 9. The 5′ ends of the sequences obtained using the four abovementioned sense primers are depicted below, with the *BamHI* restriction site being shown in italics:

native:	ATATATGGATCC ATG GAC GTC CTC CTC CTG GAG AAG GCC
Rec 1	ATATATGGATCC ATG GAT GTT TTG TTG TTG GAG AAG GCC
Rec 2	ATATATGGATCC ATG GAT GTT TTG TTG TTG GAA AAA GCT
Rec 3	ATATATGGATCC ATG GAT GTT TTG TTG TTG GAA AAA GCT
Protein:	met asp val leu leu leu glu lys ala

CTC CTG GGC CTC TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC
CTC CTG GGC CTC TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC
TTG TTG GGT TTG TTC GCC GCG GCG GTG CTG GCC ATC GCC GTC GCC
TTG TTG GGT TTG TTT GCT GCT GCT GTT TTG GCT ATT GCT GTT GCT
leu leu gly leu phe ala ala ala val leu ala ile ala val ala

AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC
AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC
AAG CTC ACC GGC AAG CGC TTC CGC CTC CCC CCT GGC CCC TCC GGC
AAA TTG ACT GGT AAA AGA TTT AGA TTG CCA CCA GGT CCA TCC GGC
lys leu thr gly lys arg phe arg leu pro pro gly pro ser gly

GCC CCC ATC GTC .....
GCC CCC ATC GTC .....
GCC CCC ATC GTC .....
GCC CCC ATC GTC .....
ala pro ile val .....

2. Transforming the yeasts

After having been digested with the restriction enzymes *BamHI* and *EcoRI*, the four above-described altered coding sequences are integrated into the vector pYeDP60, which is described by Pompon et al. (*Methods Enzymol*, 272, 1996, 51-64; WO 97/10344), the content of which is hereby incorporated by reference with regard to the plasmid, the method of insertion into the plasmid, and the method of transforming and growing the yeasts, in particular using the *Saccharomyces cerevisiae* yeast strains W(R), WAT21 and WAT11. The method for transforming and growing yeasts is also described by Pompon et al. and by Urban et al.

(*Eur. J. Biochem*, 222, 1994, page 844, 2nd column, "Yeast transformation and cell culture").

4 transformed yeast strains, designated: W73A17(native), W73A17(Rec1), W73A17(Rec2) and W73A17(Rec3), are obtained.

Example 3: Expression of CYP73A17 in the altered yeasts

The previously obtained transformed yeasts are cultured, in accordance with the method described by Urban et al. (*Eur. J. Biochem.*, 222, 1994, page 844, 2nd column, "Yeast transformation and cell culture"), in 50 ml of SGI medium at 30°C for 72 h. The cells are recovered by centrifuging at 8000 g for 10 minutes, washed with 25 ml of YPI medium, recentrifuged, and then resuspended in 250 ml of YPI medium. The cells are induced with galactose for 14-16 h, while being shaken at 160 rpm, until the cell density reaches 10^8 cells per ml. The microsomes are then prepared using the method described by Pierrel et al. (*Eur. J. Biochem.*, 224, 1994, 835-844).

The expression of CYP73A17 achieved in the case of the four strains is quantified by differential spectrophotometry using the method described by Omura and Sato (*J. Biol. Chem., 177, 678-693*). It is proportional to the number of poorly suited codons which have been altered.

The microsomal enzymic activity is measured using the method described by Durst F., Benveniste I., Schalk M. and Werck-Reichhart D. (1996) Cinnamic acid

hydroxylase activity in plant microsomes. Methods Enzymol. 272, 259-268. The results obtained after transforming WAT21 are recorded in the Table below. The activity is expressed as cinnamate 4-hydroxylase activity. The percentage additional activity (rounded values) illustrates the extent of the leap in activity which is observed after the poorly suited codons have been altered.

Strain	Activity pmol/min/µg of protein	% additional activity
W73A17 native	0.64	-
W73A17 Rec1	2.84	+340
W73A17 Rec2	4.92	+670
W73A17 Rec3	8.90	+1300

These results relating to the increase in enzymic activity confirm those relating to the increase in the expression of the protein in the yeasts. They demonstrate that alteration of the 5' end alone, even when limited (Rec1), is sufficient to obtain a very substantial improvement in the production of the enzyme by the yeast and in its enzymic activity.

Example 4: Expression of wheat CYP86A5 in the altered yeasts

The sequence encoding wheat cytochrome P450

CYP86A5, which is depicted by sequence identifier No. 10 (SEQ ID No. 10), was isolated from the wheat cDNA library described in Example 1 using the same method of operation as described for the CYP73A17 sequence and employing the complete coding sequence of *Arabidopsis thaliana* CYP86A1 as the probe. This wheat CYP86A5 sequence was altered, in accordance with the mode of operation of Example 2, using the two oligonucleotides depicted by the sequences ID No. 12 and 13 (SEQ ID No. 12 and SEQ ID No. 13) as sense and reverse primers, respectively, in order to obtain the coding sequence which is altered in accordance with the invention and which is depicted by sequence identifier No. 14 (SEQ ID No. 14).

A primer depicted by SEQ ID No. 11 was also used to enable yeasts to be transformed with the sequence encoding unmodified (native) wheat CYP86A5.

The yeasts are transformed with this new coding sequence and the expression is quantified by differential spectrophotometry in accordance with the mode of operation described in Example 2. While the natural sequence of wheat CYP86A5 is not expressed in a detectable manner, there is substantial expression in the transformed yeasts of the sequence which has been modified in accordance with the invention.

The above-described examples demonstrate unambiguously that the expression in yeasts of DNA sequences which possess a 5′ region having a high

content of codons which are poorly suited to yeasts is substantially improved when this region alone is simply recoded in accordance with the invention, ever partially, with corresponding codons which are well-suited to yeasts.